NYC Flights 2013 Analysis with R

```
library(tidyverse)
library(glue)
library(dplyr)
```

```
Warning message in data("flights"):
“data set ‘flights’ not found”
Warning message in data("airline"):
“data set ‘airline’ not found”
```

```
read.csv("flights.csv")
```

A data.frame: 336776 × 19

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tai
<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<chr>	<int>	<c
2013	1	1	517	515	2	830	819	11	UA	1545	N1
2013	1	1	533	529	4	850	830	20	UA	1714	N2
2013	1	1	542	540	2	923	850	33	AA	1141	N6
2013	1	1	544	545	-1	1004	1022	-18	B6	725	N8
2013	1	1	554	600	-6	812	837	-25	DL	461	N6
2013	1	1	554	558	-4	740	728	12	UA	1696	N3
2013	1	1	555	600	-5	913	854	19	B6	507	N5
2013	1	1	557	600	-3	709	723	-14	EV	5708	N8
2013	1	1	557	600	-3	838	846	-8	B6	79	N5
2013	1	1	558	600	-2	753	745	8	AA	301	N3
2013	1	1	558	600	-2	849	851	-2	B6	49	N7
2013	1	1	558	600	-2	853	856	-3	B6	71	N6
2013	1	1	558	600	-2	924	917	7	UA	194	N2
2013	1	1	558	600	-2	923	937	-14	UA	1124	N5
2013	1	1	559	600	-1	941	910	31	AA	707	N3
2013	1	1	559	559	0	702	706	-4	B6	1806	N7
2013	1	1	559	600	-1	854	902	-8	UA	1187	N7
2013	1	1	600	600	0	851	858	-7	B6	371	N5
2013	1	1	600	600	0	837	825	12	MQ	4650	N5
2013	1	1	601	600	1	844	850	-6	B6	343	N6
2013	1	1	602	610	-8	812	820	-8	DL	1919	N9
2013	1	1	602	605	-3	821	805	16	MQ	4401	N7
2013	1	1	606	610	-4	858	910	-12	AA	1895	N6

2013	1	1	606	610	-4	837	845	-8	DL	1743	N3

```
flights <- read.csv("flights.csv")
```

2013	1	1	611	600	11	945	931	14	UA	303	N5

```
complete.cases(flights)
sum(complete.cases(flights))
```

2013	1	1	615	615	0	1039	1100	-21	B6	709	N7
2013	1	1	615	615	0	833	842	-9	DL	575	N3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2013	9	30	2123	2125	-2	2223	2247	-24	EV	5489	N7
2013	9	30	2127	2129	-2	2314	2323	-9	EV	3833	N1
2013	9	30	2128	2130	-2	2328	2359	-31	B6	97	N8
2013	9	30	2129	2059	30	2230	2232	-2	EV	5048	N7
2013	9	30	2131	2140	-9	2225	2255	-30	MQ	3621	N8
2013	9	30	2140	2140	0	10	40	-30	AA	185	N3
2013	9	30	2142	2129	13	2250	2239	11	EV	4509	N1
2013	9	30	2145	2145	0	115	140	-25	B6	1103	N6

TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · ⋯ · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE ·
TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · TRUE · FALSE · FALSE · FALSE · FALSE · FALSE · FALSE

327346

2013	9	30	2147	2137	10	30	27	3	B6	1371	N6
2013	9	30	2149	2156	-7	2245	2308	-23	UA	523	N8
2013	9	30	2150	2159	-9	2250	2306	-16	EV	3842	N1

```
drop_na(flights)
```

2013	9	30	2203	2205	-2	2339	2331	8	EV	5311	N7
2013	9	30	2207	2140	27	2257	2250	7	MQ	3660	N5
2013	9	30	2211	2059	72	2339	2242	57	EV	4672	N1
2013	9	30	2231	2245	-14	2335	2356	-21	B6	108	N1
2013	9	30	2233	2113	80	112	30	42	UA	471	N5
2013	9	30	2235	2001	154	59	2249	130	B6	1083	N8

2013	9	30	2237	2245	-8	2345	2353	-8	B6	234	N3
2013	9	30	2240	2245	-5	2334	2351	-17	B6	1816	N3
2013	9	30	2240	2250	-10	2347	7	-20	B6	2002	N2
2013	9	30	2241	2246	-5	2345	1	-16	B6	486	N3
2013	9	30	2307	2255	12	2359	2358	1	B6	718	N5
2013	9	30	2349	2359	-10	325	350	-25	B6	745	N5
2013	9	30	NA	1842	NA	NA	2019	NA	EV	5274	N7
2013	9	30	NA	1455	NA	NA	1634	NA	9E	3393	NA
2013	9	30	NA	2200	NA	NA	2312	NA	9E	3525	NA
2013	9	30	NA	1210	NA	NA	1330	NA	MQ	3461	N5
2013	9	30	NA	1159	NA	NA	1344	NA	MQ	3572	N5
2013	9	30	NA	840	NA	NA	1020	NA	MQ	3531	N8

A data.frame: 327346 × 19

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	tai
<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<chr>	<int>	<c
2013	1	1	517	515	2	830	819	11	UA	1545	N1
2013	1	1	533	529	4	850	830	20	UA	1714	N2
2013	1	1	542	540	2	923	850	33	AA	1141	N6
2013	1	1	544	545	-1	1004	1022	-18	B6	725	N8
2013	1	1	554	600	-6	812	837	-25	DL	461	N6
2013	1	1	554	558	-4	740	728	12	UA	1696	N3
2013	1	1	555	600	-5	913	854	19	B6	507	N5
2013	1	1	557	600	-3	709	723	-14	EV	5708	N8
2013	1	1	557	600	-3	838	846	-8	B6	79	N5
2013	1	1	558	600	-2	753	745	8	AA	301	N3
2013	1	1	558	600	-2	849	851	-2	B6	49	N7
2013	1	1	558	600	-2	853	856	-3	B6	71	N6
2013	1	1	558	600	-2	924	917	7	UA	194	N2
2013	1	1	558	600	-2	923	937	-14	UA	1124	N5
2013	1	1	559	600	-1	941	910	31	AA	707	N3
2013	1	1	559	559	0	702	706	-4	B6	1806	N7
2013	1	1	559	600	-1	854	902	-8	UA	1187	N7
2013	1	1	600	600	0	851	858	-7	B6	371	N5
2013	1	1	600	600	0	837	825	12	MQ	4650	N5
2013	1	1	601	600	1	844	850	-6	B6	343	N6
2013	1	1	602	610	-8	812	820	-8	DL	1919	N9
2013	1	1	602	605	-3	821	805	16	MQ	4401	N7
2013	1	1	606	610	-4	858	910	-12	AA	1895	N6

2013	1	1	606	610	-4	837	845	-8	DL	1743	N3
2013	1	1	607	607	0	858	915	-17	UA	1077	N5
2013	1	1	608	600	8	807	735	32	MQ	3768	N9
2013	1	1	611	600	11	945	931	14	UA	303	N5
2013	1	1	613	610	3	925	921	4	B6	135	N6
2013	1	1	615	615	0	1039	1100	-21	B6	709	N7
2013	1	1	615	615	0	833	842	-9	DL	575	N3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

6 Business Question

Q1 : Top 5 flights in June 2013 ?

```
df <- drop_na(flights)
head(df, 5)
```

A data.frame: 5 × 19

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<chr>	<int>
1	2013	1	1	517	515	2	830	819	11	UA	1545
2	2013	1	1	533	529	4	850	830	20	UA	1714
3	2013	1	1	542	540	2	923	850	33	AA	1141
4	2013	1	1	544	545	-1	1004	1022	-18	B6	725
5	2013	1	1	554	600	-6	812	837	-25	DL	461

2013	9	30	2129	2059	30	2230	2232	-2	EV	5048	N7

```
top5_June_2013 <- df %>%
    filter(month == 6) %>%
    count(carrier) %>%
    arrange(desc(n)) %>%
    head(5)
```

2013	9	30	2147	2137	10	30	27	3	B6	1371	N6

```
top5_June_2013
```

2013	9	30	2150	2159	-9	2250	2306	-16	EV	3842	N1
2013	9	30	2159	1845	194	2344	2030	194	9E	3320	N9

A data.frame: 5 × 2

	carrier	n
	<chr>	<int>
1	UA	4885
2	B6	4550
3	DL	4079
4	EV	4075
5	AA	2683

2013	9	30	2203	2205	-2	2339	2331	8	EV	5311	N7
2013	9	30	2207	2140	27	2257	2250	7	MQ	3660	N5
2013	9	30	2211	2059	72	2339	2242	57	EV	4672	N1
2013	9	30	2231	2245	-14	2335	2356	-21	B6	108	N1
2013	9	30	2233	2113	80	112	30	42	UA	471	N5
2013	9	30	2235	2001	154	59	2249	130	B6	1083	N8
2013	9	30	2237	2245	-8	2345	2353	-8	B6	234	N3
2013	9	30	2240	2245	-5	2334	2351	-17	B6	1816	N3
2013	9	30	2240	2250	-10	2347	7	-20	B6	2002	N2
2013	9	30	2241	2246	-5	2345	1	-16	B6	486	N3

Q2 : Top 5 destination flights in 2013 ?

```
top5_Destination <- df %>%
    count(dest) %>%
    arrange(desc(n)) %>%
    head(5)
```

```
top5_Destination
```

A data.frame: 5 × 2

	dest	n
	<chr>	<int>
1	ATL	16837
2	ORD	16566
3	LAX	16026
4	BOS	15022
5	MCO	13967

Q3 : Which month has the lowest number of flights in the top 5?

```
flights %>%
    group_by(month)%>%
    summarise(Total_of_Flight = n()) %>%
    arrange(Total_of_Flight)%>%
    head(5)
```

A tibble: 5 × 2

month	Total_of_Flight
<int>	<int>
2	24951
1	27004
11	27268
9	27574
12	28135

Q4 : Where is the destination that people are popular to go?

```
flights %>%
    group_by(dest)%>%
    summarise(Total_of_orginal = n()) %>%
    arrange(desc(Total_of_orginal)) %>%
    head(5)
```

A tibble: 5 × 2

dest	Total_of_orginal
<chr>	<int>
ORD	17283
ATL	17215
LAX	16174
BOS	15508
MCO	14082

Q5 : Which airline had the most flights on Valentine's day?

```
install.packages("nycflights13")
library(nycflights13)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' ...
 done
```

```
flights %>%
filter(month == 2, day == 14) %>%
count(carrier) %>%
arrange(desc(n)) %>%
left_join(airlines, "carrier") %>%
select(carrier_name = name, n_flights = n) %>%
head(1)
```

A tibble: 1 × 2

carrier_name	n_flights
<chr>	<int>
United Air Lines Inc.	171

Q6 : Where were the top 3 destinations in which United Air Lines Inc arrived Valentine's day?

```
flights %>%
    filter(month == 2, day == 14, carrier == "B6") %>%
    count(dest) %>%
    arrange(desc(n)) %>%
    left_join(airports, by = c("dest" = "faa")) %>%
    select(destination = name, n)%>%
    head(3)
```

A tibble: 3 × 2

destination	n
<chr>	<int>
Fort Lauderdale Hollywood Intl	21
Orlando Intl	20
Palm Beach Intl	13